Applying IEEE 802.11e for Real-Time Services

A. Kajackas, A. Vindašius

Telecommunications Engineering Department, Vilnius Gediminas Technical University, Naugarduko str. 41, LT-03227 Vilnius, Lithuania, e-mail: algimantas.kajackas@el.vgtu.lt, antanas.vindasius@el.vgtu.lt

Introduction

Growing demand for multimedia services intend to capture major part of traffic in consumer premise networks. Wireless local area networks (WLANs) are no exception. The control over service parameters in wireless consumer access is critical, yet challenging. Transmitting miscellaneous traffic consisting of bandwidth consuming non-real-time data and time sensitive multimedia services such as voice and video requires traffic differentiation and QoS handling. However this task requires a lot of effort to gain control over shared radio channel, efficiently schedule and manage limited channel recourses.

In this paper we analyze system capacity of WLAN employing QoS enhancements. Results are compared to legacy PCF analysis, published in [1].

QoS in legacy IEEE 802.11

Initially the wireless customer access technology, based on IEEE 802.11, was designed for bursty best-effort traffic. The basic access function of the IEEE 802.11 standard, called Distributed Coordination Function (DCF), handles all customer wireless stations (STAs) independently, organizing radio channel access in a manner of contention. The principle itself suits well for low-bandwidth bursty traffic, but still brings service degradation issues, related to fairness, unpredictable delay and jitter, high collision rate, altogether leading to inefficient channel utilization.

The first attempt to adapt IEEE 802.11 based wireless access networks to delay sensitive service flows, was Point Coordination Function (PCF), which is part of the legacy IEEE 802.11 standard [2]. Basic idea of the PCF is to schedule STA transmissions in round-robin manner, issuing polls to PCF enabled STAs and following the polling list, which is made upon STA registration to access point (AP). This scheme allows introducing some level of QoS; nevertheless it lacks flexibility and versatility.

Due to low concern on QoS management at the initial stage of IEEE 802.11 technology commercialization PCF availability in commercial products of IEEE 802.11a/b/g equipment is very limited. Many vendors have chosen not to implement PCF also due to possible compatibility issues. To date, IEEE 802.11a/b/g based WLANs are considered to be only best effort access technology for basic data transmissions, unable to differentiate and provide quality enabled services.

An effort to make the difference in understanding unlicensed wireless access quality management was made by IEEE 802.11 Task Group E (TGe), which resulted in standard amendment IEEE 802.11e [3].

QoS in IEEE 802.11e

CSMA in legacy IEEE 802.11 DCF appeared to be efficient for bursty, but not predictable traffic. However, PCF suits well for predictable but not bursty transmissions. HCF in IEEE 802.11e combines both techniques making transmissions of bursty and constant bit rate traffic efficient.

Amendment features include enhanced DCF and PCF MAC mechanisms namely Enhanced Distributed Channel Access (EDCA) and HCF Controlled Channel Access (HCCA). Both of them are controlled by Hybrid Coordination Function (HCF) which resides in QoS enabled AP.

EDCA should be understood rather as relative than guaranteed QoS management, because it cannot provide strictly defined QoS parameters. The priorities in EDCA are handled in stochastic manner, providing different channel access probabilities for different packet queues. In legacy DCF, the probability to access channel is the same for all STAs, because the channel sensing and capturing mechanism provides equal rights for all contending stations. All STAs would have to sense the channel as idle for DIFS period to start transmission and perform backoff otherwise to avoid collision. The backoff time is picked randomly from interval known as Contention Window (CW) and is equal for all network nodes as well. In EDCA different Traffic Stream (TS) priorities are represented by different channel access timings called Arbitration Interframe Spacing (AIFS), AIFS, CW minimum and maximum values are different for each of four access categories (AC) or packet queues (Table 1). In order to achieve higher channel utilization efficiency, collision handling is slightly different from DCF: if virtual
collisions between different queues appear, only packet with higher priority is sent out to PHY. In case of PHY collisions, CW is not doubled as it would be in DCF, but rather incremented by fixed number called Persistent Factor (PF). Although IEEE 802.11e defines 8 separate priorities for service flows, only 4 ACs are used – the priority mapping to AC is also presented in Table 1.

<table>
<thead>
<tr>
<th>Priority</th>
<th>AC</th>
<th>AIFS</th>
<th>CWmin</th>
<th>CWmax</th>
<th>Service</th>
</tr>
</thead>
<tbody>
<tr>
<td>0, 1, 2</td>
<td>0</td>
<td>2</td>
<td>31</td>
<td>1023</td>
<td>BE</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>31</td>
<td></td>
<td>1023</td>
<td>Video</td>
</tr>
<tr>
<td>4, 5</td>
<td>2</td>
<td>1</td>
<td>31</td>
<td>63</td>
<td>Video</td>
</tr>
<tr>
<td>6, 7</td>
<td>3</td>
<td>1</td>
<td>7</td>
<td>15</td>
<td>Voice</td>
</tr>
</tbody>
</table>

Other important improvement in both contention and contention-free based transmission is mechanism allowing transmitting frame bursts. As frames in a burst are separated only by SIFS and do not require to contend for each packet in the burst, significant amount of channel bandwidth is saved thus improving channel utilization efficiency. For this matter, IEEE 802.11e amendment defines new type of timer – Transmission Opportunity (TXOP). TXOP value defines the time limit, which can be used to transmit the burst containing any number of packets as long as they fit into TXOP (including SIFSs and Acks) and belong to the same AC. In EDCA TXOP values are reported to STA trough beacon frames. In HCCA every poll packet from AP contains individual TXOP values, which are calculated by HC considering required QoS parameters and making sure that certain TXOP allocation will not unacceptably increase delay for other flows. The calculation algorithm is scheduler dependant.

![Fig. 1. EDCF AC handling (a) and prioritizing with AIFS (b)](image)

IEEE 802.11e EDCF researches [4] show that contention free burst usage in TXOP improves the global system performance at the cost of delay increase for certain traffic types. Detailed simulations [5] show quite efficient prioritizing in EDCA, however under high network load conditions low priority flows may suffer bandwidth starvation or complete blocking. In [6] EDCA is considered as not efficient for real time applications as it can not guarantee QoS even for high priority flows – under high network load collision probability is high for all service flows not looking at their priority, which result high random delays.

Voice capacity analysis of WLANs with channel access prioritizing mechanisms is presented in [7]. The analysis is based on EDCA and focuses on effect of CW. It can be seen, that in contention mode, the delays can not be fully controlled.

**Parameterized QoS with HCCA**

As specified in IEEE 802.11e, contention based access EDCA is combined with contention free access, just like in legacy IEEE 802.11 where alternating contention and contention-free periods were controlled by PCF and DCF. The control of this process is dedicated to HCF, which combines EDCA and HCCA modes.

In HCF mode each superframe consists of alternating contention-free and contention periods and starts with a beacon frame, just like PCF superframe in legacy IEEE 802.11. In contention-free period, AP polls STAs for data similarly as in PCF. QoS enabled STAs receive CF-poll frames containing TXOP allocation, which indicates burst length for the STA data to be sent towards AP.

Contention period follows right after CFP and allows STAs to contend for the channel access. This period employs EDCA mechanism, handling virtual collisions and providing priorities for 8 different packet queues. During CP the AP may capture the channel if the need to transmit CF data arises. Waiting only for PIFS it gains priority over all contending flows and grabs the channel to send CF-poll to scheduled STA. This routine when AP polls the STA during CP is called Controlled Access Period (CAP). There can be as many CAPs as AP decides to be required to meet the QoS parameters of registered TSs. Thus the main difference form legacy IEEE 802.11 is superframe structure: CP and CFP are not strictly separated, but may alternate several times in one superframe (Fig. 3).

![Fig. 3. CAP allocations in IEEE 802.11e superframe](image)

To meet required QoS parameters, AP requires to employ scheduler, which could suit for different traffic, bursty and constant flows, ensure fairness, collision free transmissions and also efficiently use spectrum. Standard proposes scheduler which is designated as reference, meaning that implementation of the scheduler scheme is implementation dependent and equipment vendors may develop their own schedulers using reference as an example.

There has been some research works on scheduler performance and scheduler proposals. Improvement proposals aim both for EDCA and HCCA, considering adoption various traffic streams. Since EDCA seems to be easy to implement and appears to be more common access
function as DCF, significant part of performance evaluations focus on EDCA operation mode only.

Contention based manner of accessing the channel in EDCA mode still results very limited QoS implementation. This drawback is also widely seen and analyzed while proposing scheduling schemes for HFC. Very few schedulers can be found for handling delay sensitive multimedia traffic. One of the attempts was J. Roy’s et al. proposal on uplink scheduler for multimedia applications, capable of attaining the QoS requirement [8]. The performance shows improvement in delay, throughput and channel utilization.

The main functions of reference scheduler include TS registration mechanism, resource admission and maintaining contention-free transmit operation in uplink and downlink.

Every STA registration procedure means registering every TS of which QoS parameters has to be considered. While registering, HC has to evaluate if the new stream may be admitted. It is required to check if it is possible to fulfill all QoS requirements the TS has demanded and if the admission of new TS will not interfere with QoS requirements of already registered Ts. Every TS is defined with the set of QoS parameters, namely Traffic Specification (TSPEC). Reference scheduler uses only mandatory TSPEC components: Mean Data Rate, Nominal MSDU Size, and Maximum Service Interval or Delay Bound.

The Service Interval (SI) has to be calculated at first. SI represents the maximum time interval between polling of specific TS. Reference scheduler uses the same SI for all admitted TSs, which is equal to minimal $T_{S\text{max}}$ value of all TSs (1).

The SI is recalculated every time new TS joins the network – in case the $T_{S\text{max}}$ of new stream is lower than current $T_{S\text{max}}$, HC updates the SI value to lower, otherwise SI remains unchanged. Also SI is adjusted to first lower submultiple of TBTT.

$$T_{S\text{max}} \leq \min(T_{S\text{max}}),$$

where $i$ – the $i$-th QoS enabled STA.

Next, the TXOP value is calculated for the new stream. Unlike SI, the TXOP is unique for every STA. TXOP is calculated only from TSPEC parameters of respective TS.

$$T_{TXOP} = \max\left(\frac{N_i \cdot L_i}{R_i} + O, \frac{M}{R_i} + O\right).$$

where $N_i$ – the count of MSDUs of $i$-th STA which fit into SI duration if transmitted at mean data rate ($\rho$); $L_i$ – nominal MSDU size of $i$-th STA; $R_i$ – physical transmission rate of the $i$-th STA; $M$ – maximum allowed MSDU size, equal to 2324 bytes [3]; $O$ – overhead expressed in time units.

Overhead includes all control packets and interframe spacings required to deliver the frame. In reference scheduler the overhead includes polling frames, PLCP and MAC overhead for data frame, ACK frames and interframe spacings (3).

$$O = t_{\text{PIFS}} + t_{\text{poll}} + t_{\text{SIFS}} + t_{\text{DATA MAC}} + t_{\text{SIFS}} + t_{\text{ACK}}$$

$N_i$ can be calculated as ceiling of the number of MSDUs that arrived at $\rho$ during the SI:

$$N_i = \left\lceil \frac{T_{S\text{max}} \cdot \rho_i}{L_i} \right\rceil.$$ (4)

The rounding up to the first integer is necessary to make sure that whole packet with overhead will fit into TXOP – as mentioned before, the frame will not be allowed to access the medium if the time required to send it out and receive ACK extends beyond TXOP limits.

When SI and TXOP are calculated, HC checks whether TS should be allowed to register. This is done by evaluating if the new TS together with already admitted TSs will not exceed beyond TBTT. The TS is admitted if the following inequality proves to be correct:

$$\frac{T_{TXOP_{i+1}}}{T_{SI}} + \sum_{i=1}^{k} \frac{T_{TXOP_{i}}}{T_{SI}} \leq \frac{T_B - T_{CP}}{T_B},$$ (5)

where $T_{TXOP_{i+1}}$ – the TXOP value of new TS; $T_{TXOP_{i}}$ – TXOP value of $i$-th admitted TS; $T_B$ – beacon interval; $T_{CP}$ – time used for EDCA traffic; $k$ – the number of already admitted TSs.

**VoIP capacity in HCCA**

Estimating channel utilization effectiveness for various traffic, usually breaks down to following types: real-time and non-real-time, constant bit rate (CBR) and variable bit rate (VBR). Since this paper is focused on handling real-time traffic applications, VBR voice conversation have been selected. Number of simultaneous VoIP calls, supported by the single wireless channel, can be a good measure of channel utilization effectiveness when using VBR delay sensitive service.

TDM-like polling schemes usually show great performance on CBR traffic, as cyclic polling ensures low latency and do not introduce significant overhead. However, VBR is more challenging, since handling bursty traffic requires adaptive scheduling in order to minimize the overhead, introduced by polling-acknowledging cycles without carrying any data.

The expected VoIP channel capacity is tightly bonded with the overhead introduced by the scheduler. The main differences in overhead comparing PCF and HCCA are the following:

1. The frame structure, allowing CAPs, is different from legacy PCF, thus latency requirements can be fulfilled even if inter-packet spacing needs to be smaller than beacon period.

2. TXOP bursting allows transmitting several packets separated by SIFS (Fig. 2), thus increasing channel utilization and decreasing the overhead.
3. Block acknowledgement scheme, introduced in IEEE 802.11e allows transmitting several packets and rather acknowledging them in blocks than each one separately.

TXOP bursting and block-ACK obviously are not helping to increase channel utilization efficiency in our case, since it is required to avoid packet grouping in real-time traffic. Thus the overhead would basically depend on frame sequences, employed by scheduler.

IEEE 802.11e amendment adds many more frame sequences to ones defined in legacy PCF and DCF routines. Those sequences include various mechanisms for polling, acknowledging, bursting, piggybacking ACKs and data to poll frames and so on. Building frame sequence in legacy PCF is straight-forward – basically only very few possible sequences are available: CF-poll, CF-poll+Data or CF-poll+CF-ACK+Data frame from AP side and CF-ACK or CF-ACK+Data frame from STA side in response to the poll. IEEE 802.11e amendment introduces much more complexity to frame sequences to achieve flexibility and higher efficiency.

General HCCA reference scheduler routine starts with HC issued CF-poll. CF-Poll contains a TXOP limit for polled STA in its QoS Control field. The CF-poll containing frame is not allowed to carry data unless aggregation subfield in the associated TSPEC is set to 1, meaning that aggregation of separate TSs is allowed.

This is reasonable, since HC polls are issued to STAs, not TSs. When separate TSs requires to be scheduled separately (no aggregation), the data from STA is carried on separate CF-ACK+QoS Data frames in TXOP frame sequence and acknowledged by AP with CF-ACK frame. Furthermore, TSs usually are defined separately for uplink (UL) and downlink (DL). Thus for single connection two TSs are required which are specified in TSPEC and scheduled independently.

Different TSs scheduling and TXOP handling creates different frame sequence scenarios; however the scope of this paper is only one connection per STA for real-time traffic. All background best-effort traffic is assumed to use EDCA.

Example of considered frame sequences is shown in Fig. 4. It shows common polling routine for UL (a) and DL (b), also packet transmission in both directions with enabled aggregation and piggybacking (c).

![Figure 4](image)

**Fig. 4.** Example frame sequences in HCCA for mutual talk state

Modeling voice transmission scenario, sending one packet per poll is most likely to happen, when SI does not exceed packetization time $T_{\text{pac}}$. In this case, TXOP will be granted for one packet only considering required overhead.

### Modeling Scenarios and Results

Studied network system is based on methodology presented in [1]. AP has a wired connection to the PSTN or IP network through a SIP proxy or H.323 gatekeeper. All calls are made from wireless nodes to outside network. The speech model is based on VBR four-state implementation according to P.59 [9]. The wireless system, based on IEEE 802.11b (PHY 11 Mbps), is also IEEE 802.11e enabled running in HCCA mode for all wireless STAs. Only reference scheduler will be evaluated in HCCA simulations. Also note that voice source and SI cycles are in perfect synchronization.

Wireless channel capacity was evaluated implementing voice source model and basic HCCA reference scheduler routines in Matlab environment.

Most common G.711 codec with voice activity detection is used for the analysis. During talk bursts codec generates 64 kbps data stream packing 20 ms voice samples into 120 B packets. After adding RTP/UDP/IP headers, we have 200 B packets being sent at 80 kbps rate. In our case we sent homogeneous voice traffic, so we will not use additional voice channels, because of TSPEC restrictions. The TS allocation algorithm controls STA registration by checking if TS with particular TSPEC can be allowed by maintaining (5) inequality.

### Table 2. TSPEC parameters for simulated TSs

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominal MSDU size ($L$)</td>
<td>200 B</td>
</tr>
<tr>
<td>Maximum MSDU Size ($M$)</td>
<td>200 B</td>
</tr>
<tr>
<td>Mean Data Rate ($\rho$)</td>
<td>80 kbps</td>
</tr>
<tr>
<td>Maximum Service Interval ($T_{\text{SIMax}}$)</td>
<td>20 ms</td>
</tr>
</tbody>
</table>

It is easy to notice, that setting $T_{\text{SIMax}}$ equal to 20 ms will force the scheduler act as TDM-like algorithm, allocating TXOPs equal to time, required to send one packet including scheduling overhead and interframe spacings. Simulations showed that using VBR, TXOPs were often unused and transmission time passed to EDCA. However, the spare channel capacity cannot be used for additional voice channels, because of TSPEC restrictions.

The TS allocation algorithm controls STA registration by checking if TS with particular TSPEC can be allowed by maintaining (5) inequality.

TS allocation algorithm with and without TS aggregation registered 16 and 11 STAs (i.e. voice channels) respectively during simulation. These numbers show the capacity limited by TS admission mechanism rather than by wireless channel. Contention free
transmission time \(T_{CFP}\), used for voice channels is presented in Fig. 5. Remaining transmission time up to \(T_{SImax}\) value is left for EDCA traffic.

Contenion period should be long enough for one maximum PPDU to transmit (6).

\[
T_{Cpmin} = t_{PPDUMax} + 2SIFS + 2T_{TimeSlot} + 8T_{ACK}.
\]

\[
T_{CFPmax} = T_{SI} - T_{Cpmin}.
\]

\(T_{Cpmin}\) value is slightly different form one calculated for original IEEE 802.11b, due to larger maximum frame body (12 bytes have been added by 11e amendment making totally 2324 bytes) and larger MAC header due to QoS related information (2 bytes have been added by 11e amendment making MAC header totally 36 Bytes length).

Additional voice channels can be gained by allowing reasonable packet loss to voice streams, i.e. granting access to more STAs than is allowed by default allocation algorithm. This hardly can be done by tuning the mandatory TSPEC parameters. Nominal or maximum MSDU size will not give any positive effect as those parameters are used for TXOP calculation and setting lower values may totally block the TS traffic. Lowering mean data rate does not make any sense either – when the system is tuned for one packet delivery per SI, TXOP value will not change even if TSPEC shows less than one packet per SI. Tuning \(T_{SImax}\) value may have harsh effect on quality and performance, because setting it lower than \(T_{pac}\) will increase the overhead and decrease channel capacity; meanwhile setting \(T_{SImax}\) higher than \(T_{pac}\) will cause packet grouping thus will increase jitter.

Capacity increase may be achieved by modifying the scheduler itself and making it suitable for VBR traffic. This was analyzed in numerous publications.

However, capacity increase can also be achieved by modifying admission control method. This way can be advantageous, because much less complexity is introduced.

The VoIP channel capacity increase, related to this kind of modification, can be seen from Fig. 6 and Fig. 7 for transmissions without and with TS aggregation respectively. The occupied transmission time distributions were obtained from simulations with modified admission control mechanism, registering as many TSs, as could be fitted in \(T_{CFPmax}\). In case TS aggregation is not used, the simulations show that no packet loss is introduced up to 15 STA, 16-th STA experience 0.45 \% packet loss, 17-th STA VoIP session shows 4.15 \% packet loss.

In case of TS aggregation, no packet loss is introduced up to 20 STA, and 21-st STA experience 0.87 \% packet loss probability.

Aggregation helps to suppress overhead, however technical expenditures of reference scheduler are higher comparing to legacy PCF when one packet per poll scenario is used for real-time services.

**Conclusions**

The simulation results show, that VoIP channel capacity can be extended by using TS aggregation and ACK piggybacking, also by modifying TS admission algorithm. Due to higher overhead of reference scheduler, the capacity is lower than legacy PCF. Yet using HCCA allows much more flexibility, since TSs can be registered with certain delay bound requirements without modifying beacon period, which is not configurable in real-time.
operation. Capacity evaluation results are summarized in Table 3.

Allocation mechanism of TGe reference scheduler is not suitable for VBR traffic, not only because polling is arranged in TDM manner, but also because of TS allocation restrictions. Simulations show, that only by modifying TS allocation algorithm, system capacity can be improved by additional 5 voice channels.

Table 3. VoIP capacity of IEEE 802.11b\textit{e} channel

<table>
<thead>
<tr>
<th>Access method (IEEE 802.11b PHY 11 Mbps)</th>
<th>Number of VoIP channels</th>
<th>Packet loss tolerance</th>
</tr>
</thead>
<tbody>
<tr>
<td>IEEE 802.11e HCCA</td>
<td>11</td>
<td>0 %</td>
</tr>
<tr>
<td>802.11e HCCA, TS aggregation, piggybacking</td>
<td>16</td>
<td>0 %</td>
</tr>
<tr>
<td>HCCA, modified admission control algorithm</td>
<td>16</td>
<td>0.45 % for 16th STA 0 % for other STAs</td>
</tr>
<tr>
<td>HCCA, TS aggregation, piggybacking, modified admission control</td>
<td>21</td>
<td>0.87 % for 21st STA 0 % for other STAs</td>
</tr>
<tr>
<td>Legacy IEEE 802.11 PCF [1]</td>
<td>26</td>
<td>0.4 % for 26th STA 0 % for other STAs</td>
</tr>
</tbody>
</table>

Downlink and uplink scheduling efficiency is another way to increase the system capacity. Most common way of implementing TS scheduling is separate for UL and DL. Therefore technological expenditures [10] for deliver single packet increase and system capacity decreases significantly. To solve this problem aggregation of TSs may be used. Aggregation bit in MAC header means that TSs within STA can be aggregated and scheduled together. Using aggregation and piggybacking additional 5 voice channels may be gained comparing to separate UL/DL scheduling. However, the aggregation makes sense only when one UL and one DL TSs are used. Thus in this scenario we propose to use EDCA for additional services.

References


Submitted for publication 2008 06 10


Analysis of IEEE 802.11e amendment and application for real time services is presented. Contention free access provided by HCCA mechanism offers QoS guarantees for latency sensitive flows. The admission algorithm is more suitable for constant bit rate traffic. The simulations show, that number of VBR VoIP sessions in wireless channel can be increased by using Traffic Stream aggregation and flexible admission control. The results are compared to ones obtained from simulations of legacy PCF. Ill. 7, bibl. 10 (in English; summaries in English, Russian and Lithuanian).


Представлен анализ стандарта IEEE 802.11e и возможности его применения для предоставления услуг реального времени. Механизм распределения доступа к среде передачи функции гибридной координации ХСС обеспечивает требуемые малые задержки. Механизм регистрации потоков ХССС приспособлен для СВР трафика, поэтому при передаче СВР трафика использованием пропускной способности канала связи может быть улучшено управление потоков и динамическим управлением процессами регистрации станций. Сравнение полученных результатов с данными ранее проведенного моделирования показывает, что ХССС механизм по количеству обслуживаемых источников уступает PCF модели. Ил. 7, библ. 10 (на английском языке; рефераты на английском, русском и литовском яз.).